

ENSEÑANZA DE ESPAÑOL COMO LENGUA EXTRANJERA A TRAVÉS DE UN SISTEMA INFORMÁTICO DE ANÁLISIS TEXTUAL DISPONIBLE EN INTERNET

Álvaro Sánchez, Socorro Bernardos y Ángel López
Universidad Politécnica de Madrid

1. INTRODUCCIÓN

La creciente importancia de los corpus de textos digitalizados para la enseñanza de lenguas ha sido puesta de manifiesto por distintos autores a lo largo de los últimos años. Tribble y Jones (1990), Flowerdew (1993), Botley *et al.* (1996) y Sánchez (2000), entre otros, han subrayado la utilidad didáctica de los corpus textuales y de los sistemas informáticos que permiten consultarlos.

El acceso gratuito por Internet a colecciones de corpus y herramientas de análisis textual ha hecho posible que tanto profesores como alumnos puedan utilizar libremente estos materiales en el aula o en casa. Sin embargo, y a diferencia de lo que ocurre en el ámbito inglés, la disponibilidad de este tipo de recursos para el estudio del español es todavía muy escasa. Actualmente sólo existen dos sistemas de referencia accesibles por Internet para la consulta de corpus españoles: el de la Real Academia Española (RAE, 2001) y el programa TACTWeb para textos literarios en castellano de la Universidad Autónoma de Madrid (TACTWeb, 2001).

Este trabajo tiene como finalidad principal la de presentar una nueva aplicación de análisis de corpus en español (ya disponible en la siguiente dirección URL: http://robinson.dia.fi.upm.es/areas/pln/analisis/Analisis_Main.html). En este nuevo sistema, cuyo nombre provisional es *Web Corpus*, destacan dos importantes novedades: en primer lugar, se ha incorporado en una única aplicación un conjunto de herramientas informáticas para el análisis textual; en segundo lugar, la interfaz gráfica ha sido especialmente diseñada para facilitar el manejo de las herramientas.

En el siguiente apartado se describirán las principales funcionalidades del sistema y, posteriormente, pasaremos a explicar algunas de sus aplicaciones pedagógicas para la enseñanza del español.

2. FUNCIONALIDADES DE LA APLICACIÓN *WEB CORPUS*: TIPOS DE ANÁLISIS TEXTUAL

Como hemos dicho, la aplicación *Web Corpus* integra diferentes herramientas de análisis de textos que pueden ser utilizadas fácilmente gracias a una interfaz gráfica muy

sencilla: un formulario dividido en campos que el usuario debe rellenar, bien tecleando directamente o bien seleccionando un elemento de una lista. La figura 1 muestra la interfaz de *Web Corpus* disponible en Internet.

A continuación explicaremos brevemente cada uno de los campos y opciones de la interfaz para comprender mejor todas las funcionalidades del sistema (los nombres de los campos se señalan en negrita).

Corpus. La aplicación, en su versión actual, contiene dos novelas españolas contemporáneas: “El Jarama” de R. Sánchez Ferlosio y “Viaje a la Alcarria” de C. J. Cela. El usuario puede elegir el corpus que prefiere consultar, seleccionando una u otra novela en la lista de este campo.

Palabra1 es el campo donde el usuario debe introducir la palabra que desea buscar en el corpus seleccionado. El usuario puede realizar consultas tanto de una palabra como de un patrón. Un patrón es un conjunto de caracteres que incluye el símbolo asterisco (*) o el de interrogación (?) como comodines de búsqueda. El asterisco representa cualquier secuencia de caracteres. Por ejemplo, el patrón “pre*” representa a todas las palabras que empiezan por “pre”. El signo de cierre de interrogación sustituye a una única letra. Por ejemplo, “amig?” representaría tanto a “amigo” como a “amiga”. Las restricciones sobre el uso de los comodines se detallan en la ayuda de *Web Corpus*. El campo **Palabra2** es idéntico a **Palabra1** y se utiliza en las consultas de coocurrencias, cuando se buscan dos palabras determinadas.

Contexto sirve para especificar el tamaño del texto en el que se mostrará la palabra buscada. El contexto de aparición de una palabra se puede medir en distintas unidades: palabra, frase (todas las palabras contenidas entre dos signos de puntuación), sentencia (palabras contenidas entre dos puntos o signos de entonación, es decir, una oración completa), carácter (o letra), y línea (70 caracteres). Para los contextos de palabra, carácter y línea se puede indicar el número de elementos que se consideran tanto a la izquierda (campo **Izquierda**) como a la derecha (campo **Derecha**) de la palabra buscada. Los valores permitidos en cada caso están especificados en la ayuda de la aplicación.

Comando es el campo que lista todos los posibles tipos de consultas que el usuario puede realizar sobre los corpus. Estas consultas se llevan a cabo gracias a las diferentes herramientas de análisis textual que incorpora *Web Corpus*. El usuario debe seleccionar un tipo de consulta en la lista de **Comando**: concordancia, coocurrencia, completa, frecuencia, adyacentes, información, sílabea, verifica. Describimos a continuación en qué consiste cada comando:

Concordancia. Busca todas las ocurrencias de una determinada palabra (la palabra o patrón del campo **Palabra1**) en el corpus elegido. Mediante el campo **Frecuencia1** (opcional) se puede determinar el número mínimo de veces que ha de aparecer esa palabra en el corpus. Los resultados se muestran al usuario en una página nueva en donde se listan todas las ocurrencias de la palabra buscada en el contexto seleccionado. La figura 2 muestra los resultados que ofrece *Web Corpus* tras buscar las concordancias de la palabra *dios* en la novela “El Jarama” en un contexto de 5 palabras a

derecha e izquierda. La figura está cortada, por lo que solo aparecen los primeros resultados encontrados (entre paréntesis, a la izquierda de cada resultado, se indica el número de línea del texto).

Coocurrencia. Esta consulta sirve para buscar las ocurrencias de dos palabras (las palabras o patrones de los campos *Palabral* y *Palabra2*) en el corpus elegido. El campo **Operador** define el tipo de coocurrencia léxica: con el operador **y** se buscan las coocurrencias de las dos palabras en el contexto especificado; con **no** se busca la *Palabral* en aquellos contextos donde no esté la *Palabra2*; con el operador **adyacentes** las dos palabras han de aparecer una al lado de otra; y con **cerca** se indica que las palabras buscadas pueden estar separadas por una distancia menor o igual a la especificada en el campo **Distancia** (que se mide en número de palabras).

Completa. Esta consulta proporciona la tabla completa de ocurrencias de una palabra, en la cual se especifica la posición de cada ocurrencia en el texto: número de palabra, de línea y de sentencia.

Frecuencia calcula el número de ocurrencias de una palabra en el texto seleccionado.

Adyacentes devuelve como resultado una tabla donde se listan todas las palabras adyacentes a la palabra buscada (es decir, aquellas palabras comprendidas entre el contexto derecho e izquierdo especificado) y el número de veces que se da la coocurrencia de las dos palabras.

Información muestra los siguientes datos del corpus seleccionado: tamaño (en bytes del fichero o corpus digital), número de líneas, número de oraciones, número de palabras distintas y número total de palabras.

Silabea realiza la división silábica de una palabra. Esta herramienta divide cualquier palabra introducida por el usuario en sus correspondientes sílabas. Por ejemplo: *aeropuerto* → *a-e-ro-puer-to*

Verifica funciona como un corrector tipo-ortográfico. Comprueba si la palabra tecleada por el usuario es o no una palabra española correcta. En caso de no serlo, devuelve una lista de palabras parecidas ordenadas por grado de similitud. Por ejemplo: *alcanze* → *incorrecta*. *Lista de palabras parecidas: 1.alcance, 2.alcana, 3.alcaide, 4.alcalde, 5.alcaná, 6.alianza, 7.altanez.*

Una vez que se hayan rellenado los campos correspondientes a una consulta, debe pulsarse el botón **Aceptar** para el sistema ejecute la petición y muestre los resultados. El número máximo de resultados que aparecen en pantalla puede fijarse con el campo **Max. Resultados**.

3. EJEMPLOS DE USO DE LA APLICACIÓN *WEB CORPUS*

En este apartado ofreceremos un pequeño repertorio de consultas a modo de ejemplos de uso de la aplicación. Estos ejemplos servirán también como muestra del enorme potencial didáctico de *Web Corpus* para el estudio lingüístico y el aprendizaje de cuestiones muy diversas de la gramática española. Todas las consultas listadas a

continuación se han realizado sobre el corpus de “El Jarama” con un contexto de 5 palabras a izquierda y derecha (invitamos al lector a probar la aplicación por Internet con estas u otras consultas diferentes).

Consultas para ejercicios de morfología derivativa y flexiva:

- Comando: *concordancia*; Palabra1: **otes* (o también: **azo*, **ones*)
- Comando: *concordancia*; Palabra1: **íais* (también: **íais*, **éis*, **eis*)

Consultas para el estudio del léxico (como variaciones de significado y acepciones de palabras polisémicas):

- Comando: *concordancia*; Palabra1: *caso*

Estudio de palabras, interjecciones y onomatopeyas en expresiones propias de la lengua coloquial:

- Comando: *concordancia*; Palabra1: *dios*
- Comando: *concordancia*; Palabra1: *chs** (también: *huy*, *oh*, *ah*, *eh*, *cá*, *mú*)

Estudios de categorías gramaticales (como por ejemplo, la distinción entre el grupo preposicional “antes de” y la conjunción “antes que”):

- Comando: *coocurrencia*; Palabra1: *antes*; Operador: *Adyacentes*; Palabra 2: *de*
- Comando: *coocurrencia*; Palabra1: *antes*; Operador: *Adyacentes*; Palabra 2: *que*

Ejercicios sobre tiempos verbales: el pretérito anterior, perífrasis verbales, la oposición entre imperfecto e indefinido, etc.

- Comando: *coocurrencia*; Palabra1: *hubo*; Operador: *Adyacentes*; Palabra 2: **ado*
- Comando: *coocurrencia*; Palabra1: *debe*; Operador: *Adyacentes*; Palabra 2: *de*
- Comando: *concordancia* Palabra1: **aba* (también: **ó*)

Construcciones y valores de la partícula “se”:

- Comando: *concordancia* Palabra1: *se*
- Comando: *coocurrencia*; Palabra1: *se*; Operador: *Adyacentes*; Palabra 2: *me*; Contexto: *palabra*; Izquierda: *0*; Derecha: *1*. (La figura 3 muestra los primeros resultados de esta consulta: las coocurrencias de las palabras “se” y “me” en el corpus de “El Jarama”).

Estudio del estilo directo y su reflejo en la lengua escrita. Para ello se pueden buscar todas las apariciones del guión en el texto. De este modo obtendremos todos los textos que contengan diálogos e intervenciones de los personajes marcadas ortográficamente (por ejemplo “Allí va – dijo Luci”). Los signos ortográficos y de puntuación que se quieran buscar deben ir entre corchetes.

- Comando: *concordancia*; Palabra1: *[-]*

Subordinación y coordinación: estudio de partículas y construcciones.

- Comando: *concordancia*; Palabra1: *cuyo*
- Comando: *coocurrencia*; Palabra1: *en*; Operador: *Adyacentes*; Palabra2: *cuanto*
- Comando: *coocurrencia*; Palabra1: *ni*; Operador: *Y*; Palabra2: *ni*
- Comando: *coocurrencia*; Palabra1: *o*; Operador: *Y*; Palabra2: *o*

Contextos de aparición de una determinada palabra. Por ejemplo, para obtener la lista de todas las palabras contiguas (sólo por la derecha) de “si” y su frecuencia de coocurrencia léxica utilizamos el siguiente comando:

- Comando: *adyacentes*; Palabra1: *si*; Contexto: *palabra*; Izquierda: *0*; Derecha: *1*.

Verificación ortográfica de plurales, acentos, diéresis y otros:

- Comando: *verifica*; Palabra1: *convoyes* (también plurales incorrectos: *albums*)
- Comando: *verifica*; Palabra1: *rehilo* (también palabras mal acentuadas: *jerseis*)
- Comando: *verifica*; Palabra1: *desagüe* (también incorrectas: *desague*)

División silábica de palabras (útil para enseñar a medir los versos o cómo partir las palabras a final de línea):

- Comando: *silabea*; Palabra1: *poeta* (también: *ruido*, *exhalar*, *desahucio*, *inspirar*...)

4. LA APLICACIÓN *WEB CORPUS* EN EL AULA DE ESPAÑOL: ESTRATEGIAS DIDÁCTICAS

Gracias a la gran variedad de consultas que permite hacer, esta aplicación puede utilizarse de muy distintas formas en la enseñanza del español como lengua extranjera. A continuación se proponen algunas estrategias didácticas útiles para la confección o realización de ejercicios lingüísticos con la ayuda de *Web Corpus*:

1. Aprendizaje inductivo / deductivo. Los resultados de las consultas a la aplicación pueden servir para que el alumno trate de inducir o descubrir a partir de ellos una determinada regla gramatical o norma de uso. Y viceversa: dada una regla como punto de partida, puede usarse la herramienta para buscar ejemplos representativos o relevantes en los corpus.
2. Niveles de estudio lingüístico. El profesor puede elaborar distintos ejercicios y clasificarlos según los niveles gramaticales que se pretendan estudiar: morfología, léxico, sintaxis, semántica, pragmática.
3. Enseñanza presencial / a distancia. La aplicación *Web Corpus*, al estar disponible en Internet, puede usarse en las clases presenciales (siempre que el aula cuente con el equipamiento preciso, al menos un ordenador con conexión a Internet), bien como complemento a las explicaciones teóricas, bien para la realización de ejercicios en el aula. *Web Corpus* también podría integrarse fácilmente en el currículum de un curso de español a distancia por Internet.

4. Enseñanza dirigida / autónoma. Las consultas y ejercicios más difíciles con la aplicación podrían estar supervisados o dirigidos por el profesor. Otros ejercicios más sencillos podrían plantearse para facilitar el aprendizaje autónomo de los alumnos.
5. Otras estrategias didácticas: realización de ejercicios individuales o en grupo; elaboración y clasificación de ejercicios según el grado de conocimientos de los alumnos (inicial, intermedio, alto); o como ayuda al profesor en la confección de distintos tipos de ejercicios (elaborar textos con lagunas, o textos para reconstruir palabras u oraciones del corpus).

5. CONCLUSIONES

Los resultados de este trabajo se resumen en dos aportaciones principales: en primer lugar, se ha presentado una nueva aplicación informática, *Web Corpus*, que permite realizar análisis de textos españoles por Internet; y en segundo lugar, se ha explicado su utilidad pedagógica para la enseñanza de español como lengua extranjera.

La aplicación *Web Corpus* es sólo un primer prototipo que admite muchas extensiones y posibles mejoras. Por ejemplo, la incorporación de nuevos corpus en español, literarios o no (ensayo, teatro, poesía, periódicos, textos científicos...) es una necesidad a corto plazo. Asimismo la utilización de corpus etiquetados permitiría realizar consultas mucho más ricas desde el punto de vista lingüístico (por ejemplo, buscar las coocurrencias de "*Lo + adjetivo*"). También están previstas mejoras en el servidor de la aplicación para optimizar los tiempos de acceso y permitir consultas simultáneas a todos los corpus. Por último, la elaboración de un repertorio completo y sistematizado de ejercicios con *Web Corpus* sería un recurso complementario de gran ayuda para profesores y alumnos.

6. REFERENCIAS BIBLIOGRÁFICAS

- Botley, S., Glass, J., McEnery, T. y Wilson, A. (1996): *Proceedings of Teaching and Language Corpora 1996*, UCREL Theoretical Papers Volume 9, Lancaster University.
- Flowerdew, J. (1993): "Concordancing as a tool in course design", *System* 21 (3), 231-243.
- RAE (2001). El sistema de consulta del CREA está disponible en la dirección <<http://www.rae.es/interno/creanet.html>> y el sistema de consulta del CORDE se encuentra en la página <<http://www.rae.es/interno/cordenet.html>>.
- Sánchez, A. (2000): "Language teaching before and after digitalized corpora. Three main issues", *Cuadernos de Filología Inglesa*, 9 (1), 5-37.
- TACTWeb (2001). El programa de la Universidad Autónoma de Madrid está accesible en la dirección <<http://mafalda.illf.uam.es/tactweb/>>.
- Tribble, C. y Jones, G. (1990): *Concordances in the classroom*, Essex, Longman.

8. Enseñanza de la gramática

